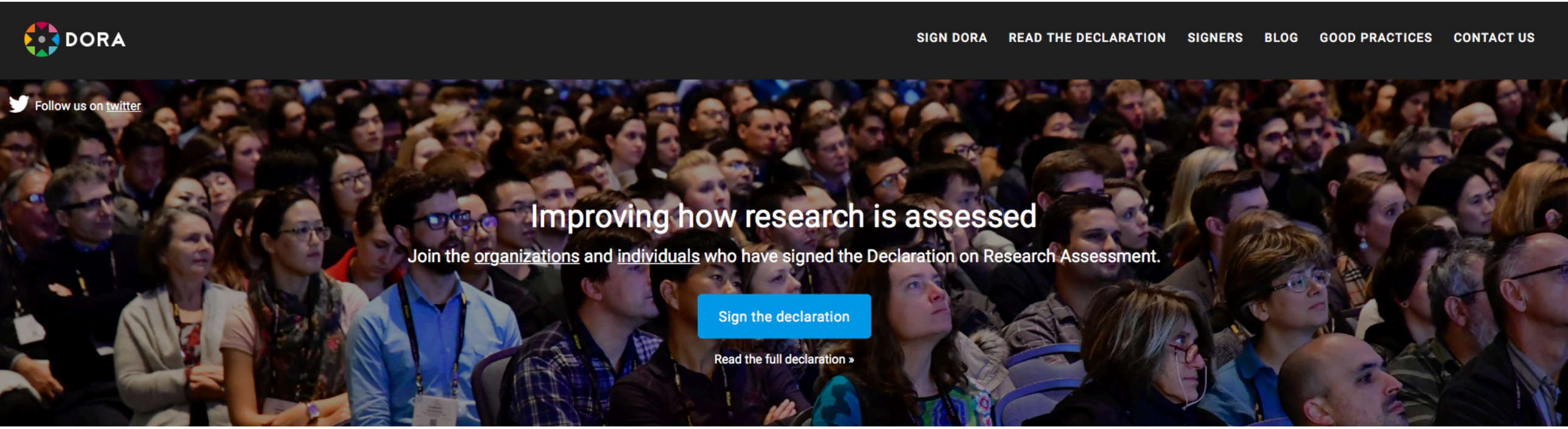
Research evaluation and DORA

Philosophy and Practice



Stephen Curry

Imperial College & DORA

UCL Open Science Day | 23 May 2019

tl;dr

Stephen Curry ✓
@Stephen_Curry

This is one of the major reasons that @DORAssessment exists – to fight for an evaluation system that values the right things.

Erick Turner
@eturnermd1

Follow

Insel: 'I spent 13 yrs at NIMH really pushing on the neuroscience and genetics of mental disorders & when I look back on that I realise that while I think I succeeded at getting lots of really cool papers published by cool scientists at fairly large costs – I think $20 billion...

> **Psychreg** @psychreg
> We Should Cure the Ills of Society to Combat the Depression Epidemic
> psychreg.org/depression-epi... @HengartnerMP

10:05 PM - 19 May 2019

27 Retweets **57** Likes

4 27 57

Tweet your reply

Erick Turner @eturnermd1 · May 19
... – I don't think we moved the needle in reducing suicide, reducing hospitalisations, improving recovery for the tens of millions of people who have mental illness.'

https://youtu.be/PeZ-U0pj9LI

tl;dr

My account

The Guardian

News | Opinion | Sport | Culture | Lifestyle

Education ▸ Schools Teachers Universities Students More

Lecturers

‘It’s cut-throat’: half of UK academics stressed and 40% thinking of leaving

Frequent rejection and a loss of control are making university staff isolated and ill, new research shows

▲ Prof Matthew Flinders says young academics can find the many rejections ‘horrible’. Photograph: Christopher Thomond/The Guardian

Anna Fazackerley

Tue 21 May 2019 07.15 BST

2,783 303

When Ed Harris, a management lecturer at a modern university, stopped sleeping and began having marriage problems, he realised he was no longer coping with the pressures of his job.

> “I felt whatever I put work into I would be blamed for what I wasn’t doing,” he says. “The system feels chaotic and you don’t understand how you’re being judged.”

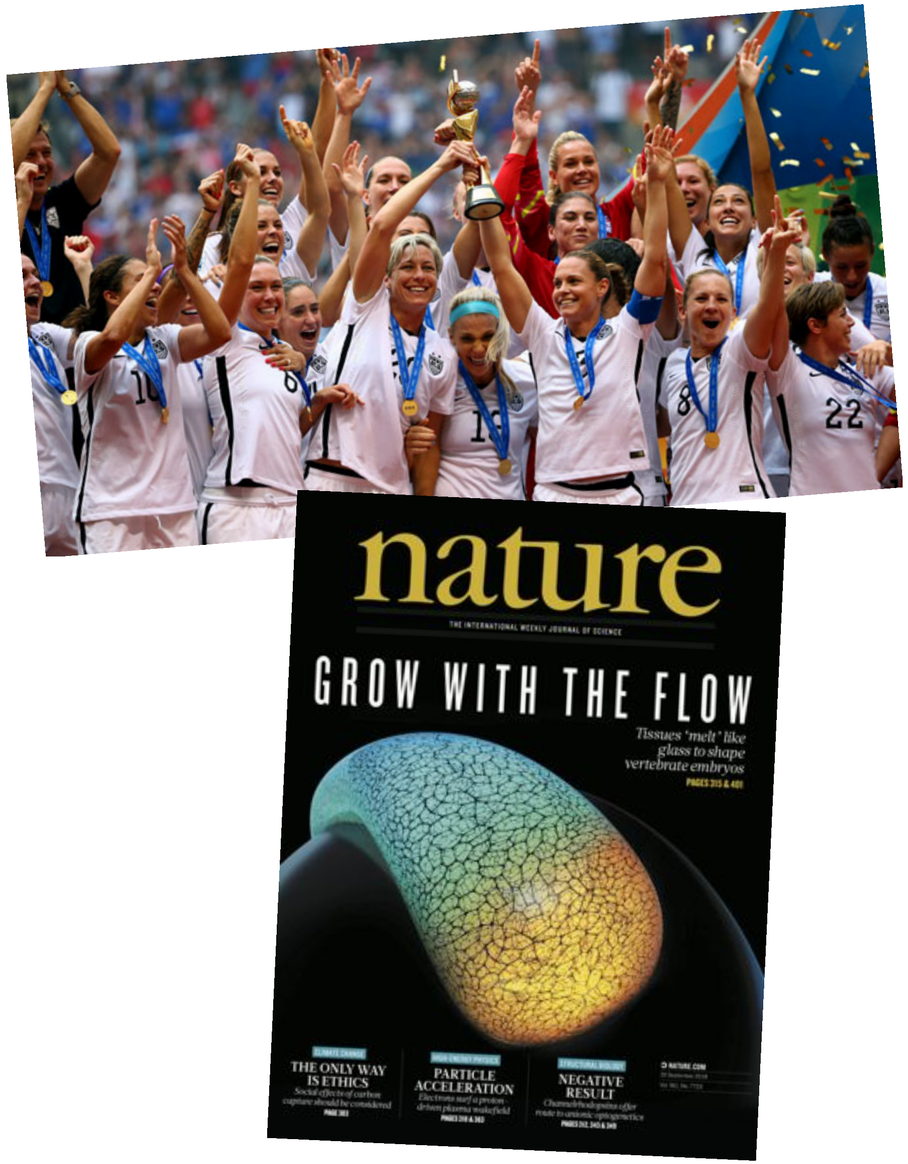
https://www.theguardian.com/education/2019/may/21/cut-throat-half-academics-stressed-thinking-leaving

Why do we need research assessment?

To invest finite (and mostly public) resources wisely

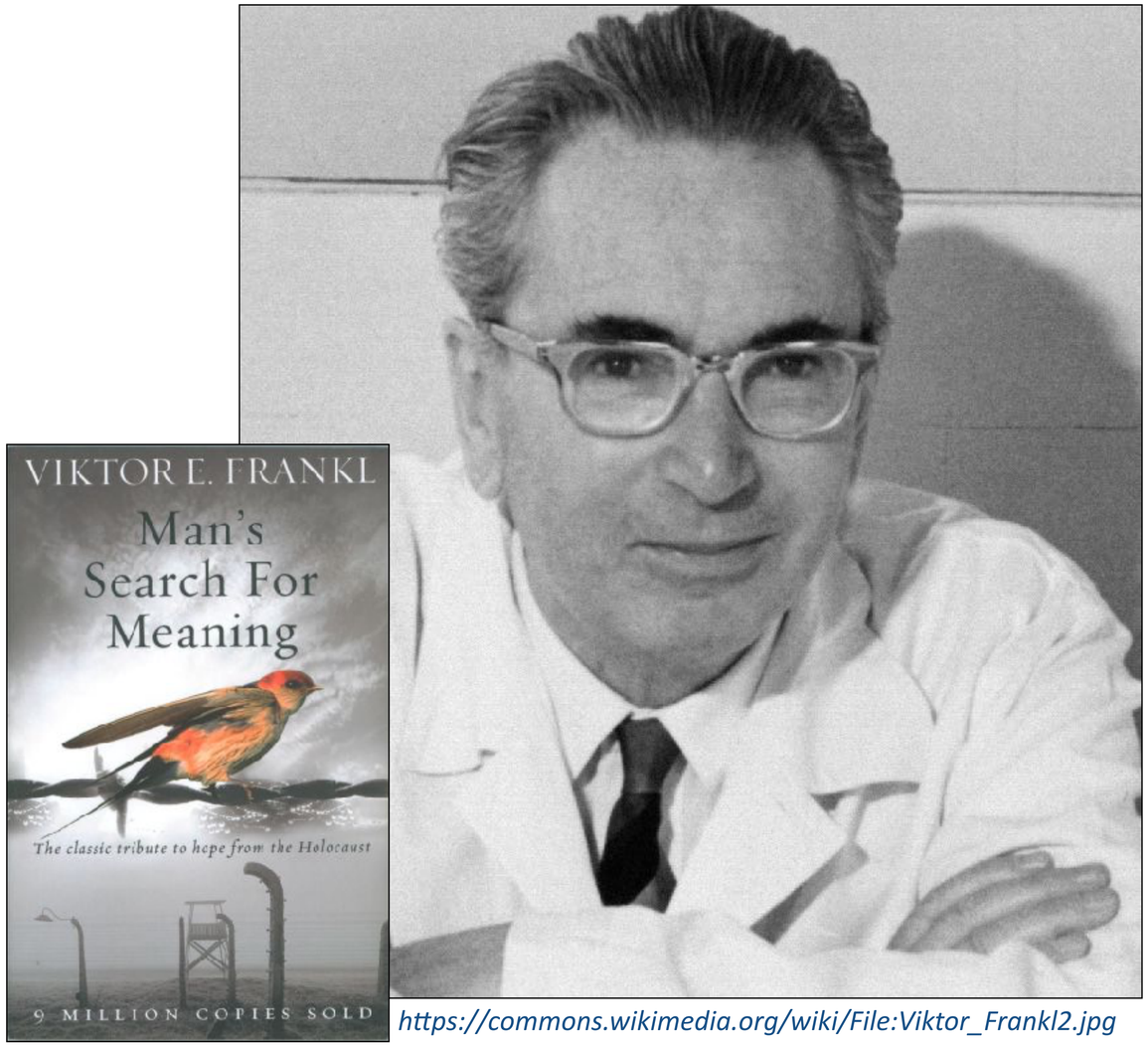
To evaluate returns on those investments

To support and encourage the best science and the best scientists



But what do we mean by 'best'?

We need to assess research but how should we define success? (Ideal world…)



https://commons.wikimedia.org/wiki/File:Viktor_Frankl2.jpg

"Don't aim at success […] for success, like happiness, cannot be pursued; it must ensue, and it only does so as the unintended side-effect of one's dedication to a cause greater than oneself…"

Viktor Frankl

We need to assess research but how should we define success? (Real world…)

Saving Science

Science isn't self-correcting, it's self-destructing. To save the enterprise, scientists must come out of the lab and into the real world.

Daniel Sarewitz

"much of the problem can be traced back to a **bald-faced but beautiful lie** upon which rests the political and cultural power of science. […] It goes like this:

> *Scientific progress on a broad front results from the **free play of free intellects**, working on **subjects of their own choice**, in the manner dictated by their **curiosity** for exploration of the unknown."*

Sarewitz's article and responses – https://www.thenewatlantis.com/publications/must-science-be-useful

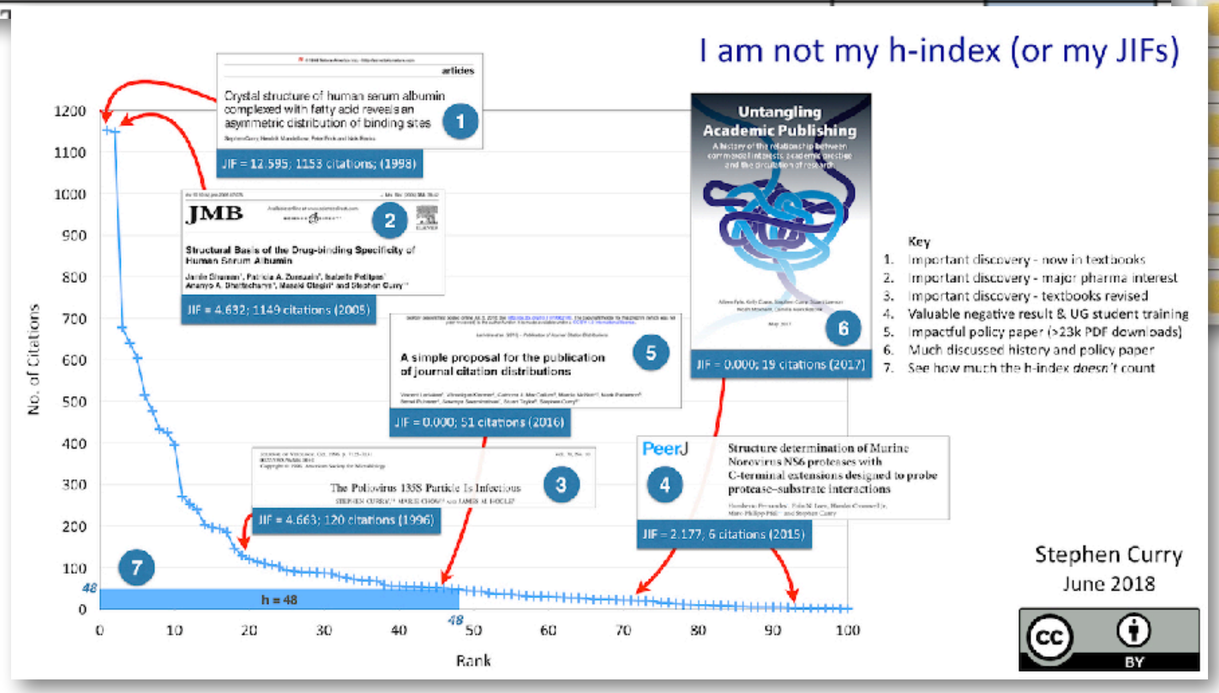
Measurement has its uses…

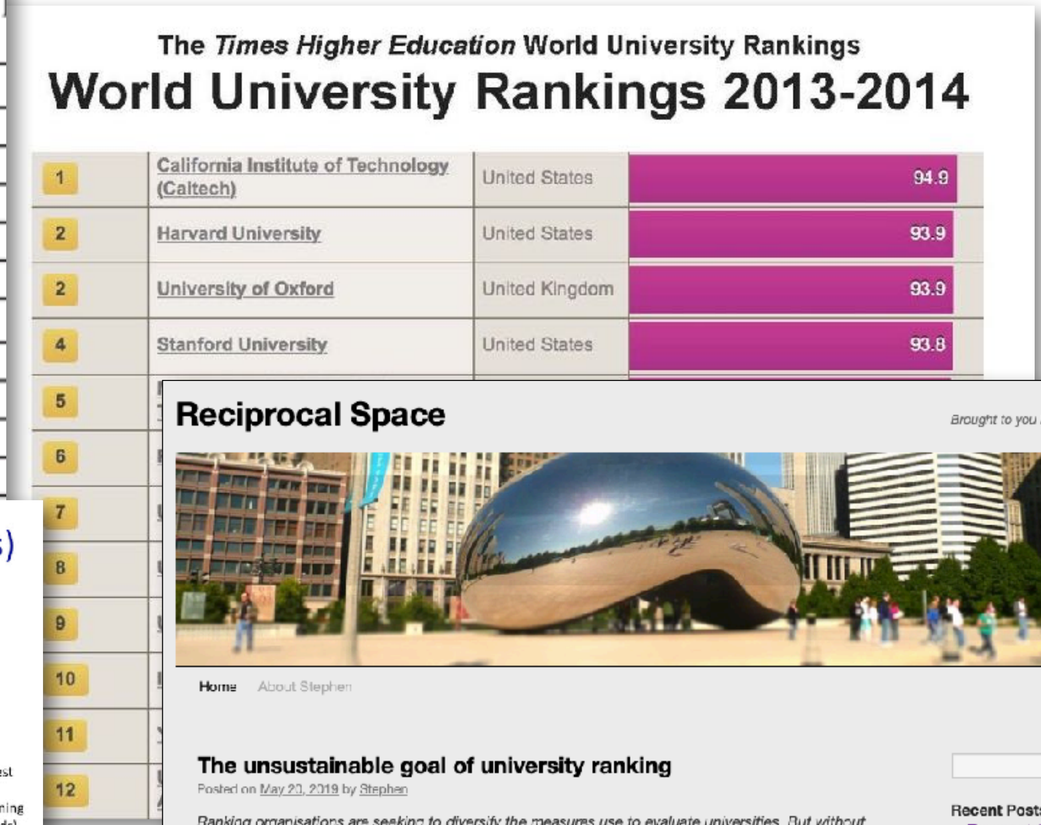
http://www.aronline.co.uk/blogs/news/news-uk-car-manufacturing-enjoys-bumper-2013/

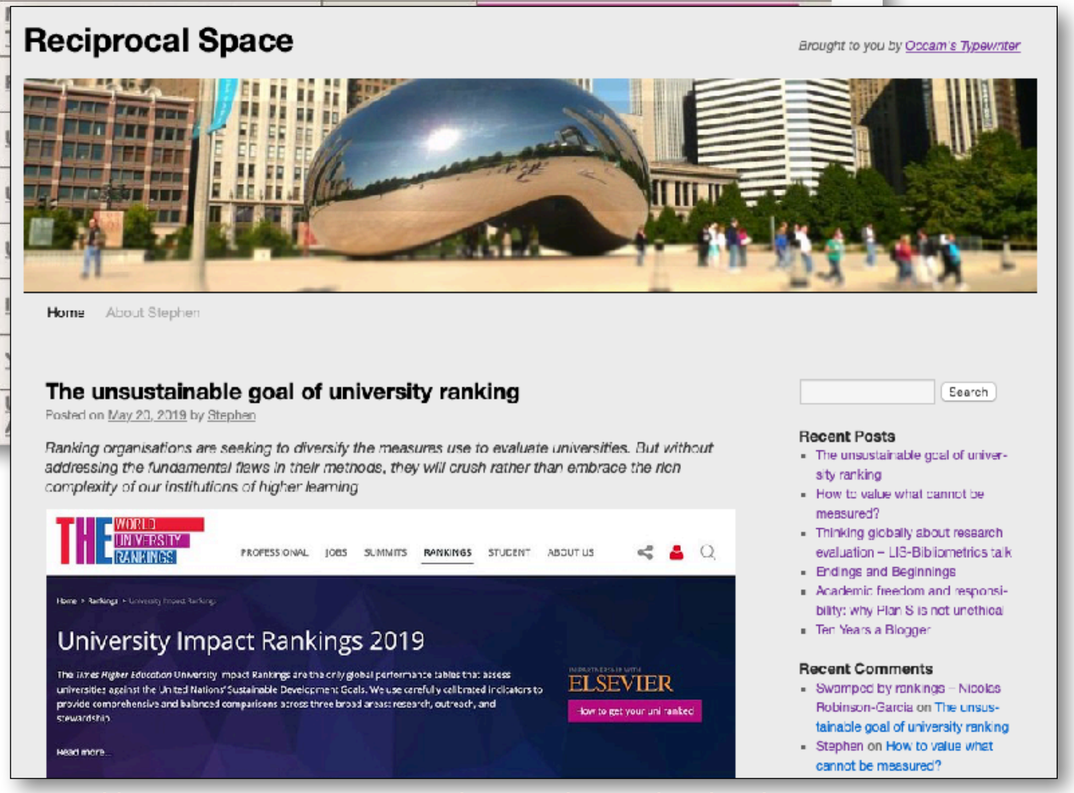
https://www.nuh.com.sg/patients-and-visitors/patients-and-visitors-guide/choice-of-accomodation/ward-types.html

…but where are the limits?

Rank	Full Journal Title	Total Cites	Journal Impact
1	CA-A CANCER JOURNAL FOR CLINICIANS	28,839	244.585
2	NEW ENGLAND JOURNAL OF MEDICINE	332,830	79.258
3	LANCET	233,269	53.254
4	CHEMICAL REVIEWS	174,920	52.613
5	Nature Reviews Materials	3,218	51.941
6	NATURE REVIEWS DRUG DISCOVERY	31,312	50.167
7	JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	148,774	47.661
8	Nature Energy	5,072	46.859
9	NATURE REVIEWS CANCER	50,407	42.784
10	NATURE REVIEWS IMMUNOLOGY	39,215	41.982
11	NATURE	710,766	41.577







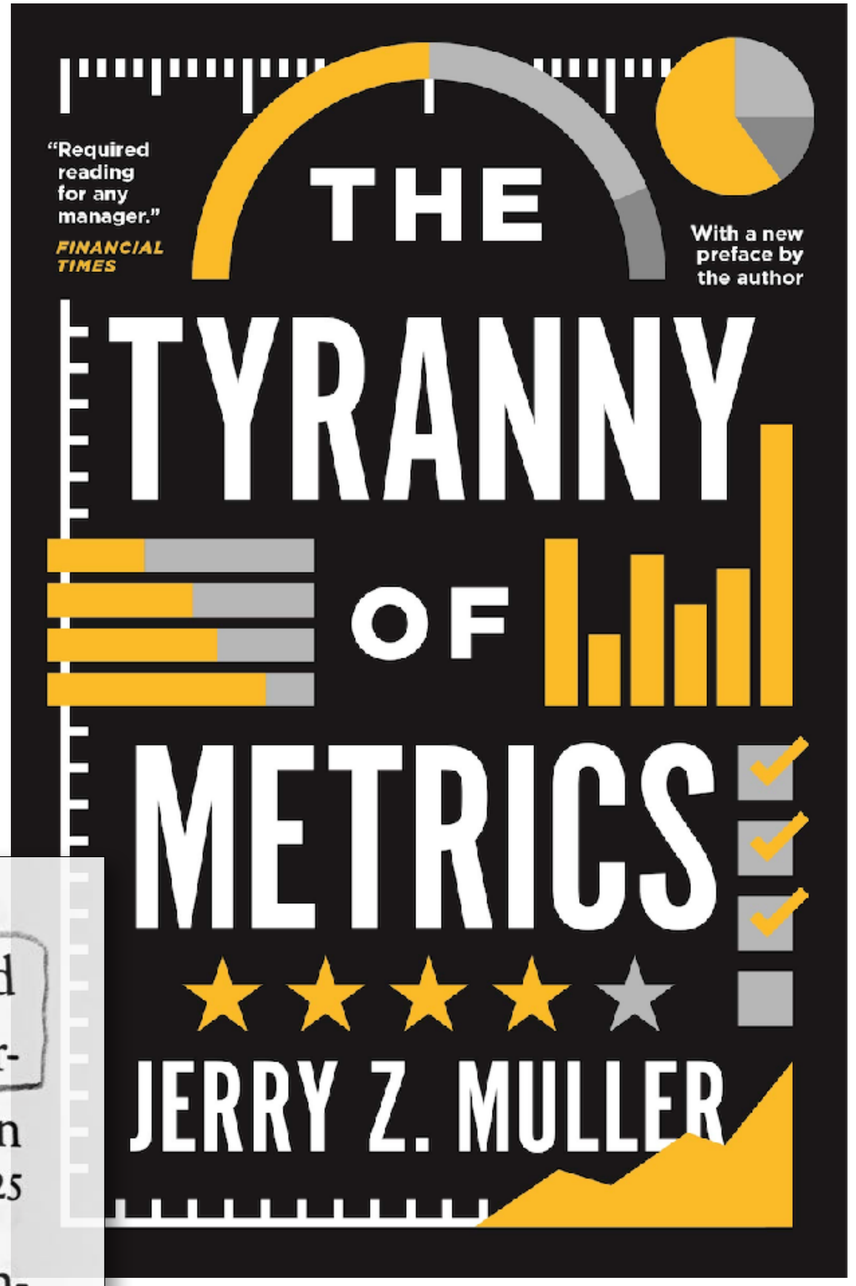
http://occamstypewriter.org/scurry/2019/05/20/unsustainable-goal-university-ranking/

…and who gets to decide when to use metrics?

The new managerialism - generalists, not generals…

Metrics can be useful but:

- not if misapplied by people who do not understand their context
- not if tied too tightly to extrinsic rewards



onto them—known in the trade as "window dressing."[24]

A focus on measurable performance indicators can lead managers to neglect tasks for which no clear measures of performance are available, as the organizational scholars Nelson Repenning and Rebecca Henderson have recently noted.[25] Unable to count intangible assets such as reputation, em-

Negative effects of over-reliance on metrics based on academic outputs

Sick of Impact Factors

Posted on August 13, 2012 by Stephen

I am sick of impact factors and so is science.

The impact factor might have started out as a good idea, but its time has come and gone. Conceived by Eugene Garfield in the 1970s as a useful tool for research libraries to judge the relative merits of journals when allocating their subscription budgets, the impact factor is calculated annually as the mean number of citations to articles published in any given journal in the two preceding years.

http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors/

"I'm really excited. We just had a big paper in Cell… !"

Postdoc (University of Y)

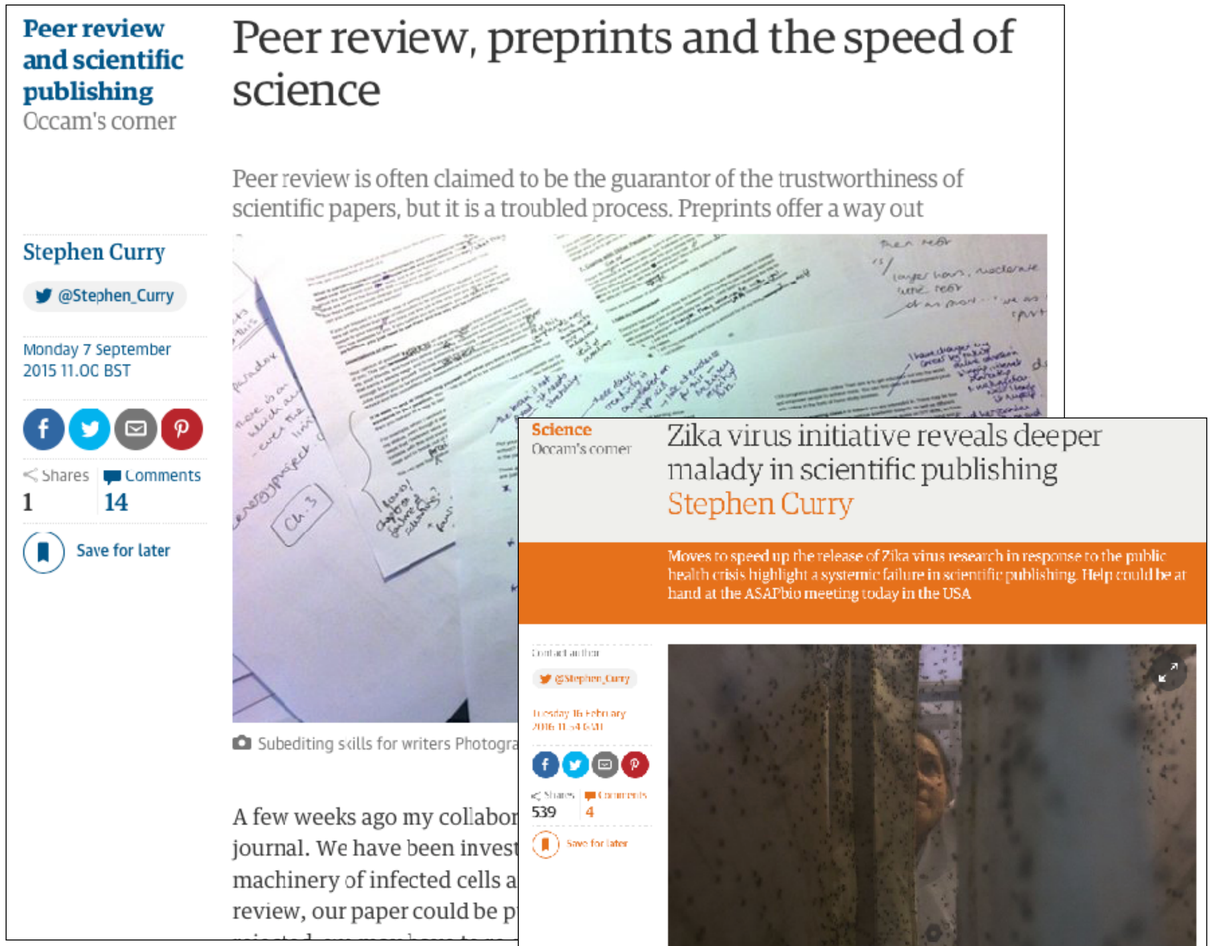
"*Despite* personal ideals and good intentions, in this incentive and reward system **researchers find themselves pursuing not the work that benefits public or preventive health or patient care the most, but work that gives most academic credit** and is better for career advancement."

Frank Miedema

https://blogs.bmj.com/openscience/2018/01/24/setting-the-agenda-who-are-we-answering-to/

- slows publication & reduces productivity
- positive bias in the literature
- JIF correlates with retraction rate
- impact on reliability & public trust?
- devaluation of other important activities
- stress on the individual

Accentuate the positive: how *open* science can be *better* science



Peer review and scientific publishing
Occam's corner

Peer review, preprints and the speed of science

Peer review is often claimed to be the guarantor of the trustworthiness of scientific papers, but it is a troubled process. Preprints offer a way out

Stephen Curry

@Stephen_Curry

Monday 7 September 2015 11.00 BST

Shares 1 Comments 14

Save for later

Subediting skills for writers Photogra

A few weeks ago my collabor
journal. We have been invest
machinery of infected cells a
review, our paper could be p

Science
Occam's corner

Zika virus initiative reveals deeper malady in scientific publishing

Stephen Curry

Moves to speed up the release of Zika virus research in response to the public health crisis highlight a systemic failure in scientific publishing. Help could be at hand at the ASAPbio meeting today in the USA

@Stephen_Curry

Tuesday 16 February 2016 11.54 GMT

Shares 539 Comments 4

Save for later

Too far behind a screen - Zika scientists are set to benefit from the rapid release of research on the virus
Photograph: Victor Moriyama/Getty Images

In response to the rapid spread of Zika virus across Central and South America, now declared to be an international public health emergency by the World Health Organisation, a consortium of research funders, institutes and publishers have committed to sharing data and results relevant to the crisis "as rapidly and openly as possible."

Preprints: faster communication; worldwide access

Focus on the content, not the container (journal)
- Valuable groundwork for journal-independent evaluation

Largest possible audience (sharing + scrutiny = reliability)
- Same applies to **OA papers**

Practice encourages **open peer review**

Data sharing: scrutiny benefits (reliability)

Better for changing the world (utility & impact; *e.g.* Zika crisis)

DORA: the declaration

One general recommendation:

Do not use journal-based metrics, such as Journal Impact Factors, as a **surrogate measure** of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

17 positive recommendations for different stakeholders:

- funders
- **institutions**
- publishers
- data providers
- researchers

For institutions:

4. **Be explicit** about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the **scientific content of a paper is much more important than publication metrics** or the identity of the journal in which it was published.

5. For the purposes of research assessment, **consider** the value and impact of **all research outputs** (including datasets and software) in addition to research publications, and **consider a broad range of impact measures** including qualitative indicators of research impact, such as influence on policy and practice.

https://sfdora.org/read/

DORA: the *campaign*

San Francisco Declaration on Research Assessment

- 6 years old; >14,000 individuals and >1300 organisations signed
- 2018: New funding, new steering group, new URL - *sfdora.org*
- New Roadmap for **action**:
  - Increase awareness of the need to develop alternatives to the JIF
  - **Research and promote best practice in research assessment.**
  - Extend the global and disciplinary impact of DORA
- New international advisory board – a truly global initiative

https://www.nature.com/articles/d41586-018-01642-w

WORLD VIEW *A personal take on events*

DAVE GUTTRIDGE

Words were a good start — now it is time for action

Five years ago, the Declaration on Research Assessment was a rallying point. It must now become a tool for fair evaluation, urges **Stephen Curry**.

Declarations are bound to fall short. The 240-year-old United States Declaration of Independence holds it self-evident that "all men [*sic*] are created equal", but equality remains a far-off dream for many Americans.

The San Francisco Declaration on Research Assessment (DORA; https://sfdora.org) is much younger, but similarly idealistic. Conceived by a group of journal editors and publishers at a meeting of the American Society for Cell Biology (ASCB) in December 2012, it proclaims a pressing need to improve how scientific research is evaluated, and asks scientists, funders, institutions and publishers to forswear using journal impact factors (JIFs) to judge individual researchers.

DORA's aim is a world in which the content of a research paper matters more than the impact factor of the journal in which it appears. Thousands of individuals and hundreds of research organizations now agree and have signed up. Momentum is building, particularly in the United Kingdom, where the number of university signatories has trebled in the past two years. This week, all seven UK research councils announced their support.

Impact factors were never meant to be a metric for individual papers, let alone individual people. They're an average of the skewed distribution of citations accumulated by papers in a given journal over two years. Not only do these averages hide huge variations between papers in the same journal, but citations are imperfect measures of quality and influence. High-impact-factor journals may publish a lot of top-notch science, but we should not outsource evaluation of individual researchers and their outputs to seductive journal metrics.

Most agree that yoking career rewards to JIFs is distorting science. Yet the practice seems impossible to root out. In China, for example, many universities pay impact-factor-related bonuses, inspired by unwritten norms of the West. Scientists in parts of Eastern Europe cling to impact factors as a crude bulwark against cronyism. More worryingly, processes for JIF-free assessment have yet to gain credibility even at some institutions that have signed DORA. Stories percolate of research managers demanding high impact factors. Job and grant applicants feel that they can't compete unless they publish in prominent journals. All are fearful of shrugging off the familiar harness.

So, DORA's job now is to accelerate the change it called for. I feel the need for change whenever I meet postdocs. Their curiosity about the world and determination to improve it burns bright. But their desires to pursue the most fascinating and most impactful questions are subverted by our systems of evaluation. As they apply for their first permanent positions, they are already calculating how to manoeuvre within the JIF-dependent managerialism of modern science.

There have been many calls for something better, including the Leiden Manifesto and the UK report 'The Metric Tide', both released in 2015. Like DORA, these have changed the tenor of discussions around researcher assessment and paved the way for change.

It is time to shift from making declarations to finding solutions. With the support of the ASCB, Cancer Research UK, the European Molecular Biology Organization, the biomedical funder the Wellcome Trust and the publishers the Company of Biologists, *eLife*, F1000, Hindawi and PLOS, DORA has hired a full-time community manager and revamped its steering committee, which I head. We are committed to getting on with the job.

Our goal is to discover and disseminate examples of good practice, and to boost the profile of assessment reform. We will do that at conferences and in online discussions; we will also establish regional nodes across the world, run by volunteers who will work to identify and address local issues.

IT'S WORTH DOING THE EXPERIMENT TO PROPERLY EVALUATE EVALUATION.

This week, for example, DORA is participating in a workshop at which the Forum for Responsible Metrics — an expert group established following the release of 'The Metric Tide' — will present results of the first UK-wide survey of research assessment. This will bring broader exposure to what universities are thinking and doing, and put the spotlight on instances of good and bad practice.

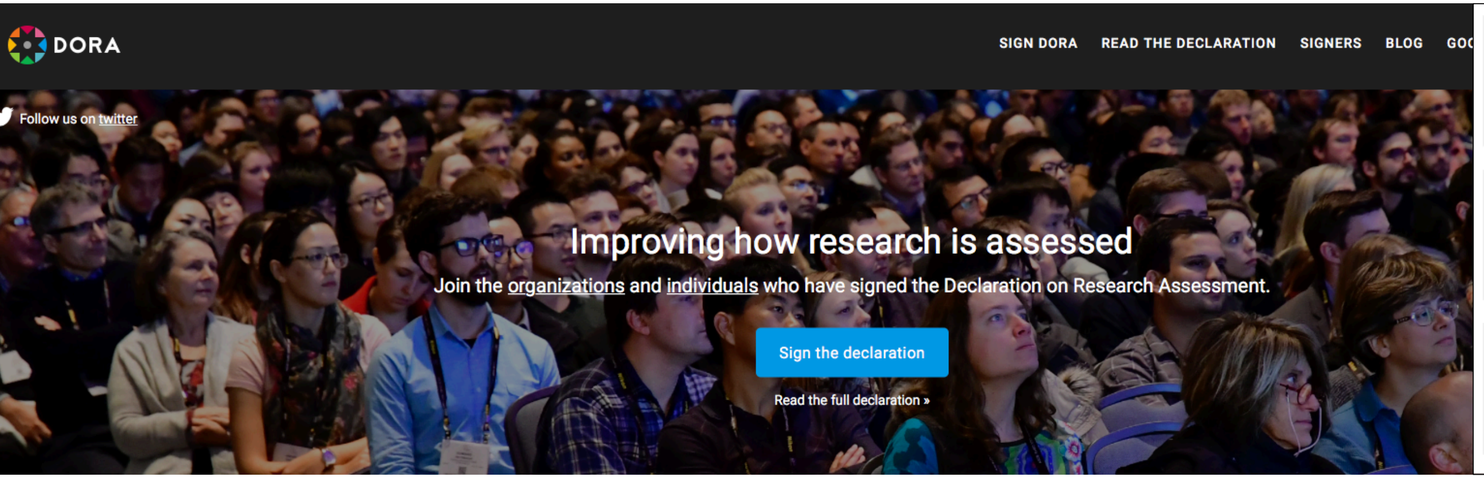
We have to get beyond complaining, to find robust, efficient and bias-free assessment methods. Right now, there are few compelling options. I favour concise one- or two-page 'bio-sketches', similar to those rolled out in 2016 by the University Medical Centre Utrecht in the Netherlands. These let researchers summarize their most important research contributions, plus mentoring, societal engagement and other valuable activities. This approach could have flaws. Perhaps it gives too much leeway for 'spin.' But, as scientists, surely we can agree that it's worth doing the experiment to properly evaluate evaluation.

This is hard stuff: we need frank discussions that grind through details, with researchers themselves, to find out what works and to forestall problems. We need to be mindful of the damage wrought to the careers of women and minorities by bias in peer review and in subjective evaluations. And we need to join in with parallel moves towards open research, data and code sharing, and the proper recognition of scientific reproducibility.

Declarations such as DORA are important; credible alternatives to the status quo are more so. True success will mean every institution, everywhere in the world, bragging about the quality of their research-assessment procedures, rather than the size of their impact factors. ■

Stephen Curry *is a professor of structural biology and assistant provost for equality, diversity and inclusion at Imperial College London. He is also chair of the DORA steering group.*
e-mail: s.curry@imperial.ac.uk

8 FEBRUARY 2018 | VOL 554 | NATURE | 147

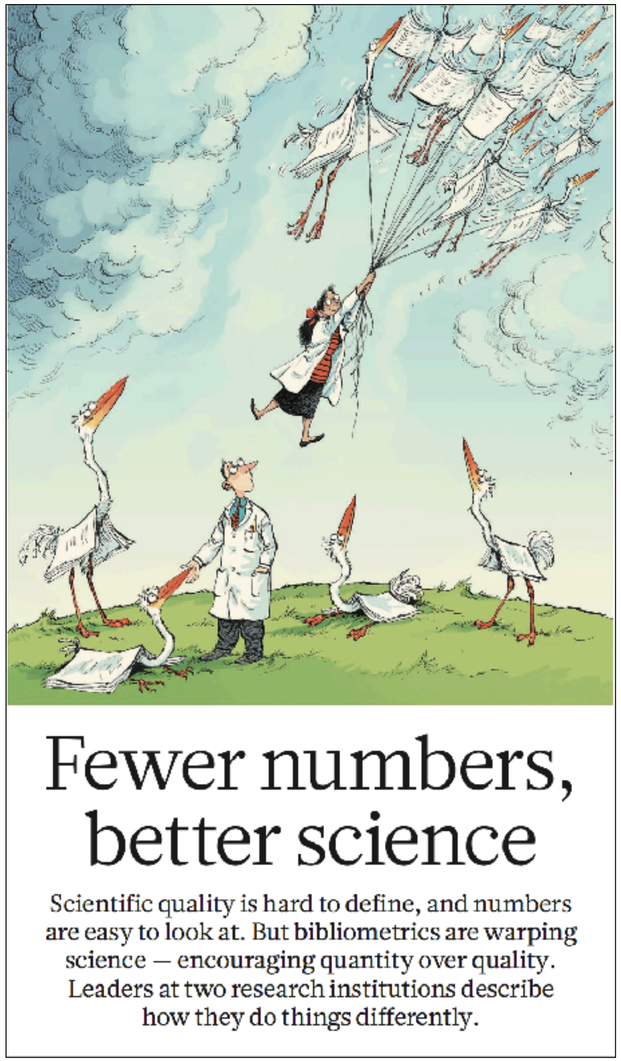


ascb an international forum for cell biology™ | CANCER RESEARCH UK | The Company of Biologists

eLIFE | EMBO excellence in life sciences | F1000Research Open for Science

Hindawi | PLOS | wellcome

New tools and processes for assessment



Fewer numbers, better science

Scientific quality is hard to define, and numbers are easy to look at. But bibliometrics are warping science — encouraging quantity over quality. Leaders at two research institutions describe how they do things differently.

http://www.nature.com/news/fewer-numbers-better-science-1.20858

Researcher assessment at UMC Utrecht

1. Research, publications, grants
2. Managerial & academic duties
3. Mentoring & teaching
4. Clinical work (if applicable)
5. Entrepreneurship & community outreach

Charité University Hospital, Berlin

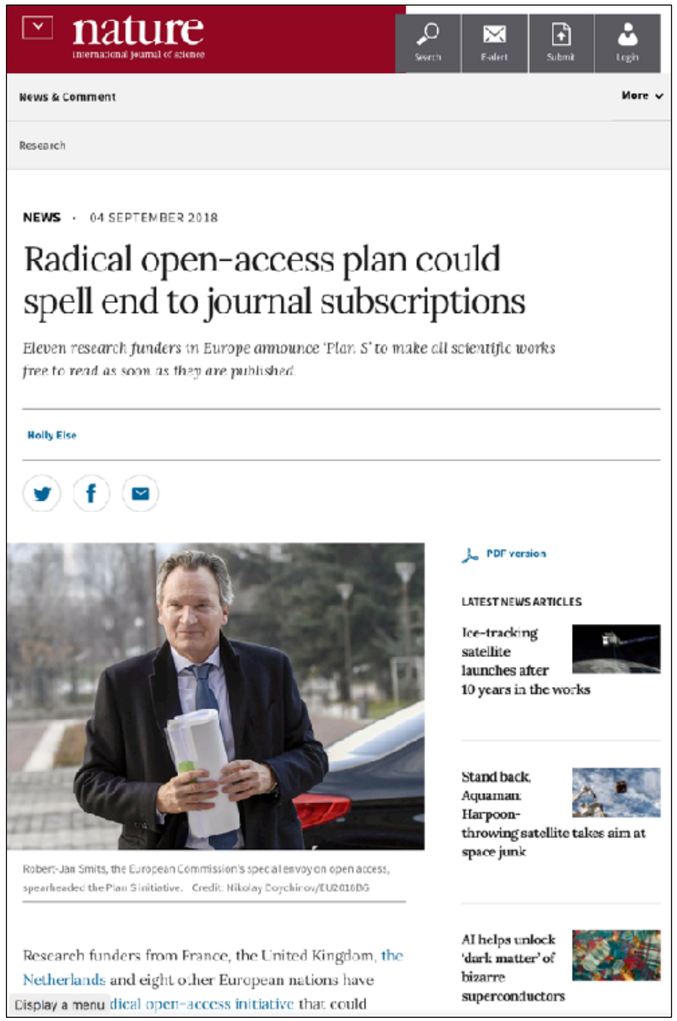
- Scientific contribution to your field
- Your 5 most important papers
- Contribution to open science
- Your most important collaborations

New tools and processes for assessment

DORA session at ASCB |EMBO (Dec 2018)

DORA session at AAAS (Feb 2019)

More info & ideas at: *https://sfdora.org/*



https://www.nature.com/articles/d41586-018-06178-7

"We also understand that researchers may be driven to do so by a misdirected reward system which puts emphasis on the wrong indicators (e.g. journal impact factor). We therefore commit to fundamentally revise the incentive and reward system of science, using the San Francisco Declaration on Research Assessment (DORA) as a starting point.

https://www.scienceeurope.org/coalition-s/

We need to assess research but how should we define success?

ANNALS OF MEDICINE JULY 29, 2013 ISSUE

SLOW IDEAS

Some innovations spread fast. How do you speed the ones that don't?

By Atul Gawande

Why do some innovations spread so swiftly and others so slowly? Consider the very different trajectories of surgical anesthesia and antiseptics, both of which were discovered in the nineteenth century. The first public demonstration of anesthesia was in 1846. The Boston surgeon Henry Jacob Bigelow was approached by a local dentist named William Morton, who insisted that he had found a gas that could render patients insensible to the pain of surgery. That was a dramatic claim. In those days, even a minor tooth extraction was excruciating. Without effective pain control, surgeons learned to work with slashing speed. Attendants pinned patients down as they screamed and thrashed, until they fainted from the agony. Nothing ever tried had made much difference. Nonetheless, Bigelow agreed to let Morton demonstrate his claim.

We yearn for frictionless, technological solutions. But people talking to people is still the way that norms and standards change.

ILLUSTRATION BY [illegible]

http://www.newyorker.com/magazine/2013/07/29/slow-ideas

"We yearn for frictionless, technological solutions. But people talking to people is still how the world's standards change."

Atul Gawande

What should success look like?

Reliable, rapidly communicated, accessible, high-quality **research** that transforms our understanding of the world and can change it for the better.

Researchers who collaborate, who feel a duty of care to group members & colleagues, and a responsibility to the societies of which they are an integral part.

A **research system** that values the people within it, that considers their quality of life, their mental health, and that provides the training and processes to seek out the creative vigour of diversity.

Thank you

s.curry@imperial.ac.uk
@Stephen_Curry